

PROYECTO DE POLÍTICAS EN SALUD Y EDUCACIÓN

Manual de Instalación y configuración

Proceso de ETL y Transferencia de datos para el sistema de Información Hospitalaria, InfHos

Diciembre 2013

Proceso de Extracción, Transformación, Carga y Transferencia de datos desde el Sistema InfHos hacia la base de datos de reportes y la construcción de cuadros de mando



USAID
DEL PUEBLO DE LOS ESTADOS UNIDOS DE AMÉRICA

PROYECTO DE POLÍTICAS EN SALUD Y EDUCACIÓN

Tabla de contenidos

1. Introducción

El Sistema de Información Hospitalaria (InfHos) está capturando información de la producción de los hospitales donde se encuentra implementado durante algunos años. Dicha información será utilizada para generar cuadros de mando.

Se ha establecido un proceso automatizado para transferir dicha información a una base de datos local para reportes y paralelamente a un Data Warehouse para la alimentación a cuadros de mando.

Este proceso, requieren una serie de programas y ambientes para instalar y configurar, dicho software está en el siguiente listado.

1. Oracle Java Development Kit (JDK) 7.u45 o superior
2. Pentaho Data Integrator Community Edition (PDI CE – KETTLE) 4.9
3. ActiveMQ 5.9.0
4. Proceso ejecutable diseñado para infhos

Cada uno de los programas es indispensable para el funcionamiento del proceso.

Este manual contiene los pasos para obtener, instalar, configurar y ejecutar cada paquete de software y código prediseñado, con los parámetros necesarios para su funcionamiento ideal.

2. Descarga de recursos

2.1. Oracle JDK 7u45 o superior

Este es el Kit de desarrollo de Java en su versión 7u45 o superior. Se puede descargar el instalador para Windows desde:

http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html

En el enlace se encuentran 2 instaladores en su versión para Windows:

1. jdk-7u45-windows-x64.exe para arquitecturas de 64 bits
2. jdk-7u45-windows-i586.exe para arquitecturas de 32 bits

Se deben aceptar los términos y condiciones y descargar la versión que corresponda (125 MB aproximadamente).

En el WEBSERVER. al finalizar la descarga únicamente se debe ejecutar el instalador y seguir las instrucciones de la instalación normal. Luego de la instalación se recomienda reiniciar el equipo.

2.2. ActiveMQ 5.9.0 o superior

Este es un sistema de colas de mensajes, hace las veces de repositorio de información en forma de mensajes entre los diferentes sistemas de forma remota.

Se puede descargar su paquete binario desde:

http://www.apache.org/dyn/closer.cgi?path=/activemq/apache-activemq/5.9.0/apache-activemq-5.9.0-bin.zip

En el WEBSERVER, al finalizar la descargar, se debe descomprimir en el directorio raíz de la unidad C: del sistema operativo.

2.3. Pentaho Data Integrator 4.4.0 (KETTLE)

La herramienta de integración de datos de Pentaho, también conocida como PDI o KETTLE, tiene la finalidad de crear un trabajo programado para la extracción, trasformación y carga (proceso ETL por sus siglas en ingles) de los datos del sistema INFHOS hacia la base de datos de reportes internos.



Para los trabajos ETL ya programados se debe usar la versión 4.4.0 y esta se puede descargar desde:

http://sourceforge.net/projects/pentaho/files/Data%20Integration/4.4.0-stable/pdi-ce-4.4.0-stable.zip/download

Debe alojarse en el WEBSERVER en la ruta C:\ZV\

2.4. Repositorio de recursos desarrollados

Ya existen recursos desarrollados específicamente para el funcionamiento del proceso y son necesarios para interconectar cada herramienta.

Se puede descargar el archivo comprimido desde:

http://hpp-gt.org/infhos/recursos.zip

y debe alojarse en el WEBSERVER en la ruta C:\ZV\

3. Configuración de ambientes

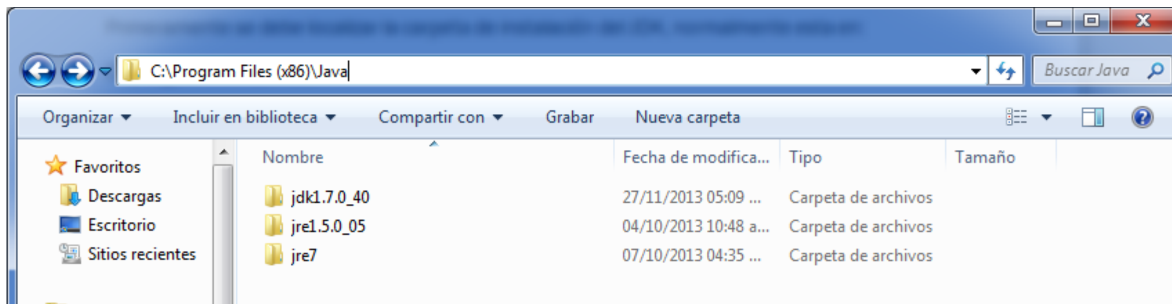
3.1. Variables de entorno

Se deben configurar las variables del sistema para las rutas de acceso a la instalación del JDK en el WEBSERVER

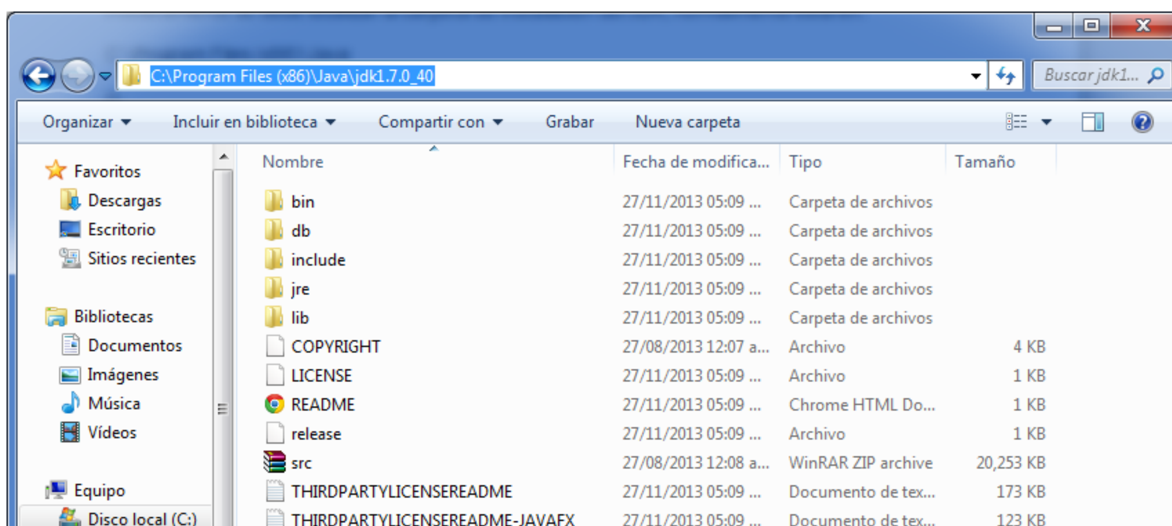
Primeramente se debe localizar la carpeta de instalación del JDK, normalmente esta en:

C:\Program Files (x86)\Java O C:\Program Files\Java

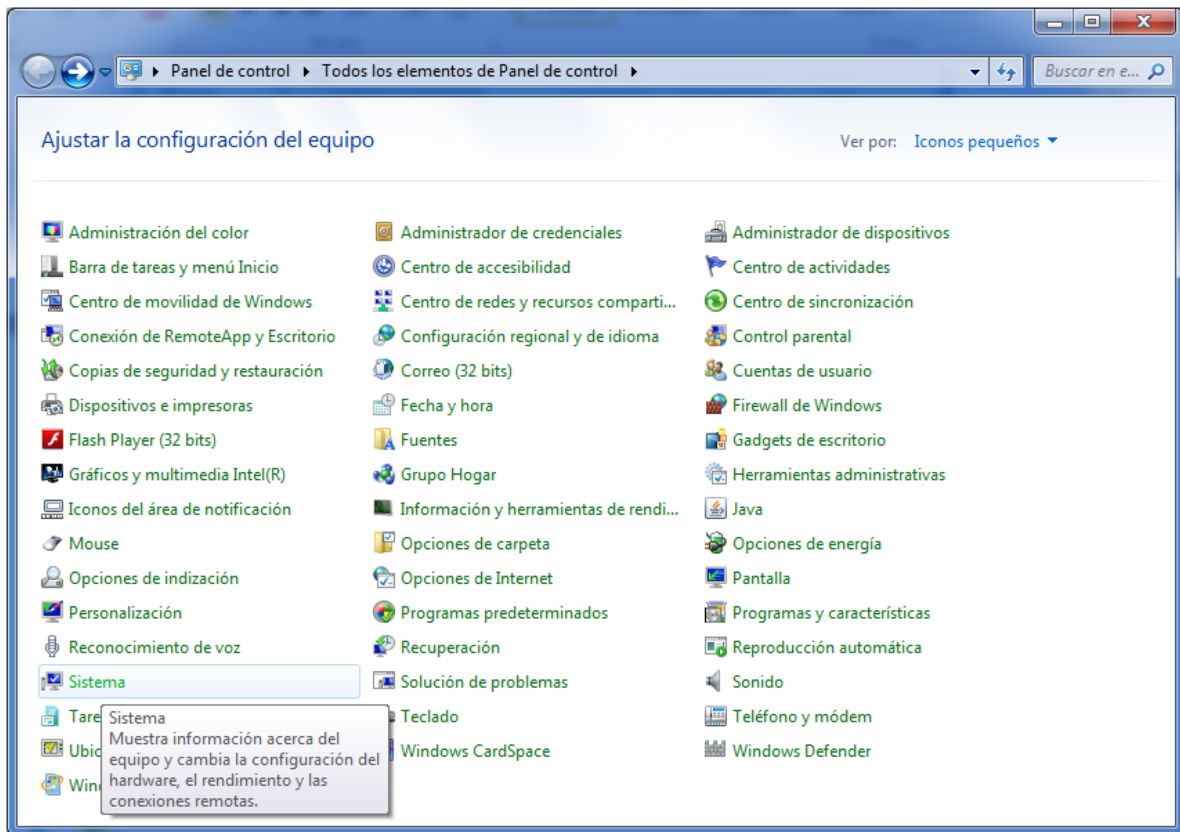
Se deben encontrar el archivo JDK como lo muestra la siguiente imagen



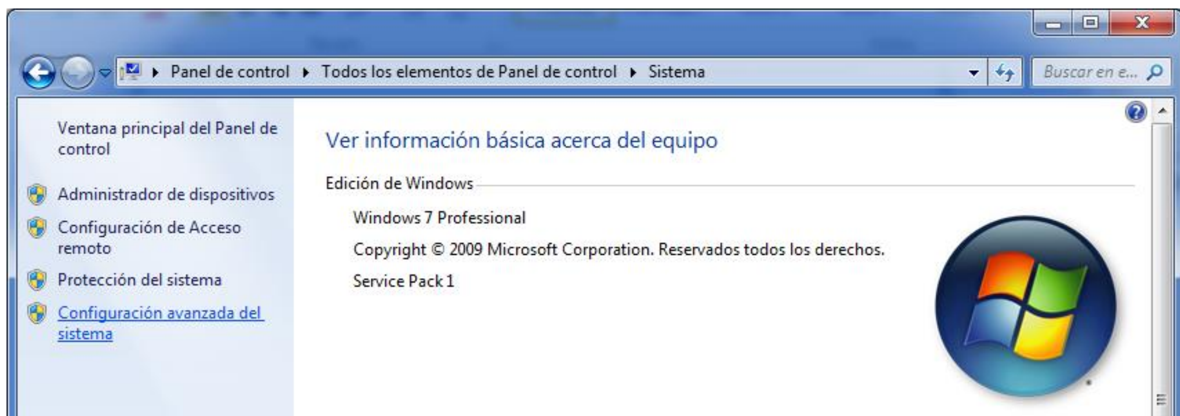
Luego ingresar a la carpeta JDK y copiar la ruta de acceso



Abrir el panel de control de Windows y buscar el icono "Sistema" o "System" si el lenguaje del sistema operativo esta en español

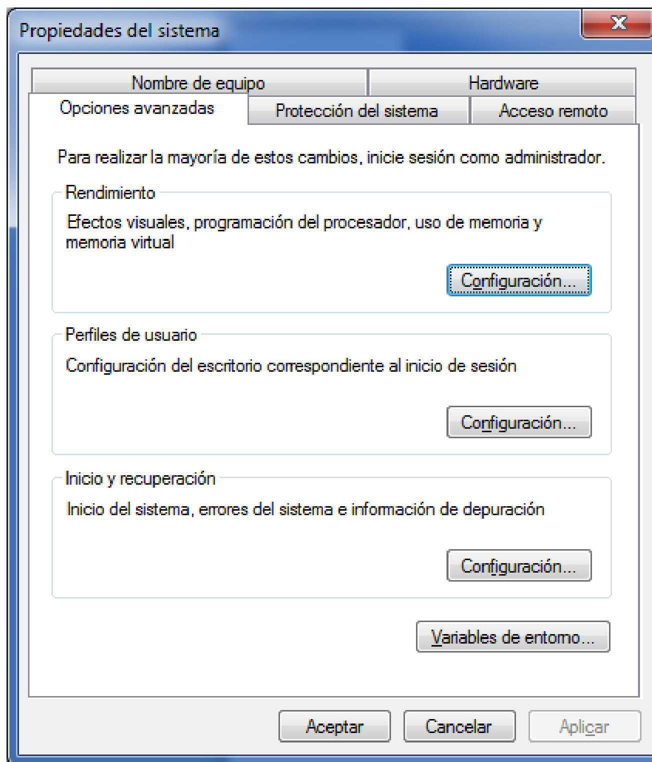


Luego buscamos en el menú del lado izquierdo el enlace hacia "Configuración avanzada del sistema"

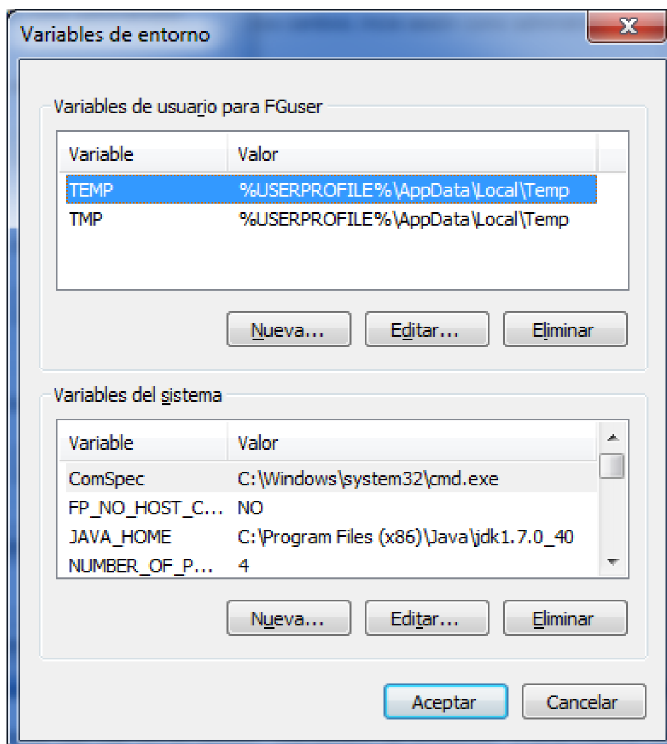


Esto desplegará una ventana emergente con varias pestañas y opciones

En la pestaña "opciones avanzadas" hay 3 secciones y un botón al final que dice "Variables de entorno"



Se debe dar clic en el botón "Variables de entorno" y una nueva ventana emergente aparecerá:



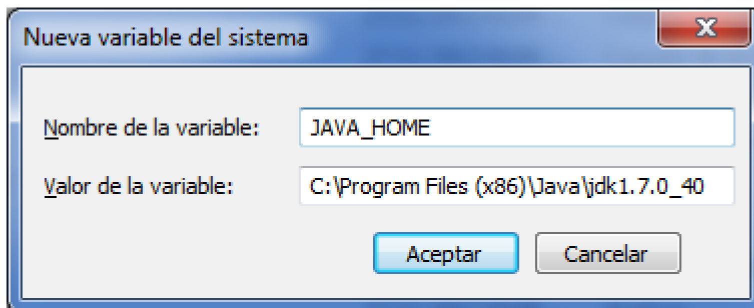
En esta ventana existen 2 secciones, la primera en la parte superior de la ventana, contiene todas las variables de entorno definidas exclusivamente para las sesiones del usuario actual. Y la parta inferior muestra las variables de entorno global; es decir que afectan a todo el sistema.

Si se cuenta con un usuario administrador del sistema operativo lo aconsejable es agregar una nueva variable a las variables globales.

Dando clic al botón “Nueva” aparecerá la ventana para agregar una variable:

En el campo “nombre de la variable” se debe escribir JAVA_HOME

Y en el campo de “Valor de la variable” se debe ingresar el path hacia la carpeta de instalación del JDK



Luego vamos dando clic en aceptar hasta que cierren todas las ventanas

3.1.1. Comprobación

Para comprobar si efectivamente Java está instalado y tenemos configurada la variable de entorno puedes escribir en una ventana de comandos de Windows (Command Prompt):

Ver instalación de java

java -version

Comprobar la variable de entorno

Echo %JAVA_HOME%

Estos comandos deben retornar los mensajes como los muestra la imagen siguiente:

```
C:\Windows\system32\cmd.exe

C:\Users\FGuser>java -version
java version "1.7.0_45"
Java(TM) SE Runtime Environment (build 1.7.0_45-b18)
Java HotSpot(TM) 64-Bit Server VM (build 24.45-b08, mixed mode)

C:\Users\FGuser>echo %JAVA_HOME%
C:\Program Files (x86)\Java\jdk1.7.0_40
```

3.2. ActiveMQ como servicio de Windows

El paquete comprimido de ActiveMQ se debe descomprimir en el directorio C:\ZV\ del WEBSERVER

Dentro de la carpeta “apache-activemq-5.9.0-bin” se debe ingresar a la ruta *apache-activemq-5.9.0\conf* y renombrar el archivo *activemq.xml* por activemq_origin.xml y luego descargar sobre esta misma carpeta el archivo *activemq.xml* pre-configurado para crear el puente entre el ActiveMQ local y su contraparte remoto en el Data Warehouse. El archivo se puede descargar desde:

http://hpp-gt.org/infhos/amq/descargar-config.php

Para instalar el servicio de activeMQ como un servicio de Windows, se debe acceder a la carpeta donde se descomprimió el paquete y luego a la carpeta ***bin*** en función de la arquitectura con que cuente el servidor ingresar a win32 o win64 respectivamente.

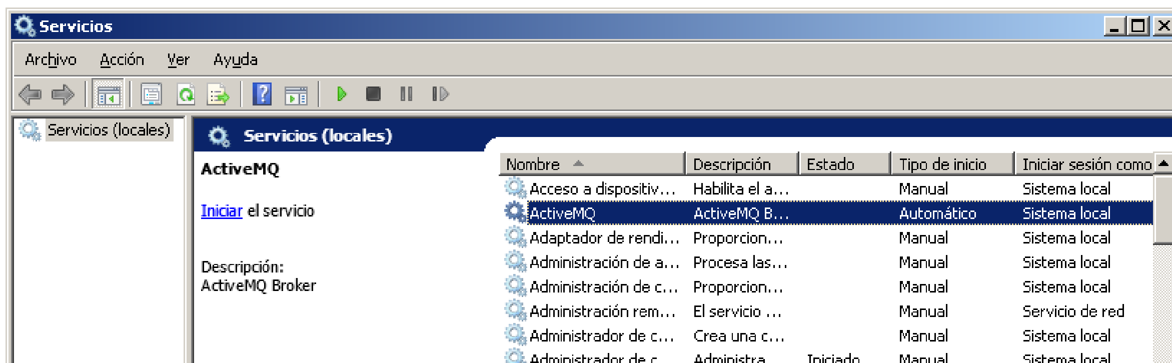
Dentro de estas carpetas se encuentra un archivo ejecutable *InstallService.bat*, hacer clic derecho sobre el mismo y de nuevo clic en ejecutar como administrador.

Esto instalara activeMQ como servicio de Windows con los parámetros necesarios para realizar el puente en las colas.

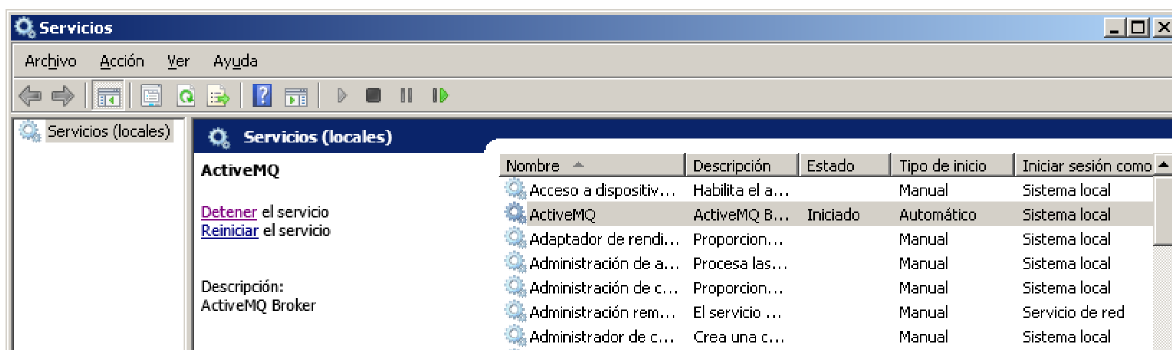
Para iniciar el servicio se debe acceder al administrador de servicios del sistema operativo, la ruta a seguir es:

Inicio –> Panel de Control -> Herramientas administrativas -> Servicios

Se debe identificar el servicio de ActiveMQ, seleccionarlo con un clic y dar de nuevo clic en el enlace “iniciar” al lado izquierdo de la ventana.



En ese momento el servicio debe quedar iniciado



Consola Web administrativa de activeMQ

Para ingresar a la consola web administrativa del sistema se debe acceder en el navegador a la url:

http://localhost:8161/admin

Los datos de acceso son:

Usuario: admin

Clave: admin

Consola administrativa:

ActiveMQ™

Home | Queues | Topics | Subscribers | Connections | Network | Scheduled | Send

Welcome!

Welcome to the Apache ActiveMQ Console of **localhost** (ID:T410s-52838-1386949704702-0:1)

You can find more information about Apache ActiveMQ on the Apache ActiveMQ Site

Broker

Name	**localhost**
Version	**5.9.0**
ID	**ID:T410s-52838-1386949704702-0:1**
Uptime	**36.573 seconds**
Store percent used	**0**
Memory percent used	**0**
Temp percent used	**0**

Para verificar las colas de mensajes se debe ingresar al menú "Queues" y aparecerá la lista de colas.

ActiveMQ™

Home | Queues | Topics | Subscribers | Connections | Network | Scheduled | Send

Queue Name [ ] Create

Queues

Name	Number Of Pending Messages	Number Of Consumers	Messages Enqueued	Messages Dequeued	Views	Operations
infhos	0	1	0	0	Browse Active Consumers atom rss	Send To Purge Delete

Si aún no se han enviado datos, esta lista aparecerá vacía.

Si está todo configurado correctamente, en la columna de "Number of Consumers" de la fila para la cola "infhos" aparecerá un "1".

4. Bases de datos

4.1. Repositorio para KETTLE

La herramienta de integración de datos de Pentaho, KETTLE, necesita un repositorio para poder alojar las configuraciones de cada trabajo y transformación especifica.

Por este motivo es necesario crear una base de datos en el SERVER con nombre ***KETTLE*** y con accesos:

1. nombre de usuario: ***Xaman***
2. contraseña: ***Xaman***

4.2. Creación de la base de datos de dbReportes

Todos los datos transformados se alojan en una base de datos distinta a la transaccional.

El procedimiento es el siguiente:

1. Ejecutar consulta del archivo *Create Database dbReportes.sql* para crear la base de datos dbReportes.
2. Seleccionar "dbReportes" como la base de datos de uso para ejecutar consultas, esto mediante el comando "USE [dbReportes];"
3. Ejecutar consulta del archivo *2013-11-11 63 alter table tbl_produccion.sql* para crear la tabla tbl_produccion.
4. Ejecutar consulta del archivo *2013-11-11 64 create trigger tr_Modificar.sql* Ejecutar consulta del archivo *2013-11-11 65 create trigger tr_Insertar.sql*

5. Proceso de Extracción, Transformación y Carga (ETL)

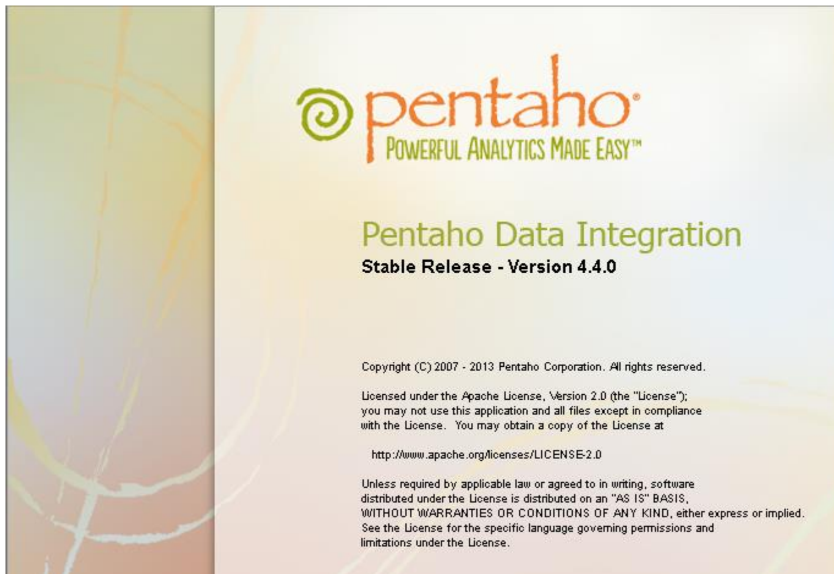
Se configurará un proceso para extraer los datos de la base de datos de producción de InfHos y trasladarlos a otra base de datos (dbReportes) para ponerlos a disposición para la generación de reportes y transferencia para la construcción de cuadros de mando.

Para esto se deben integrar el trabajo y transformación pre-diseñadas para este fin.

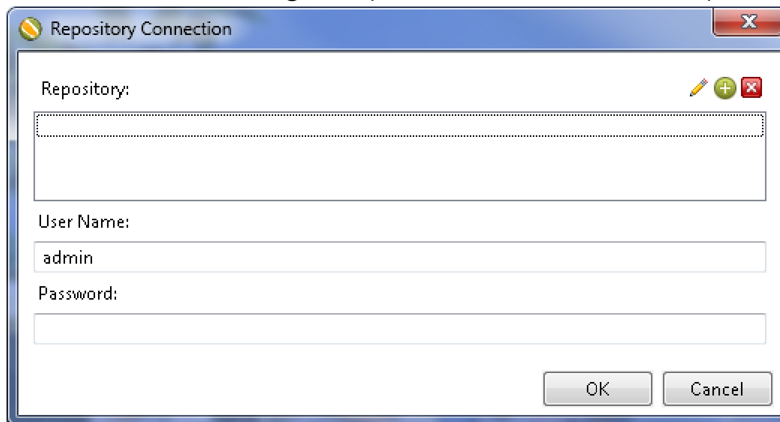
5.1. Configuración de repositorio local

Se debe descomprimir el paquete descargado *pdi-ce-4.4.0-estable.zip* e ingresar la carpeta descomprimida *pdi-ce-4.4.0-estable*. Dentro de la misma ingresar a la carpeta *data-integration* y se debe dar doble clic sobre el archivo *spoon.bat*

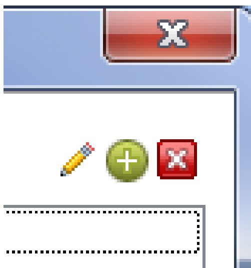
Aparecerá esta ventana:



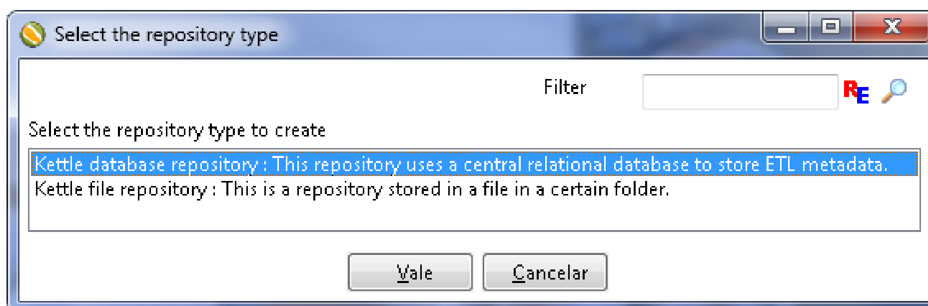
Al finalizar de cargar aparecerá la ventana para seleccionar un repositorio:



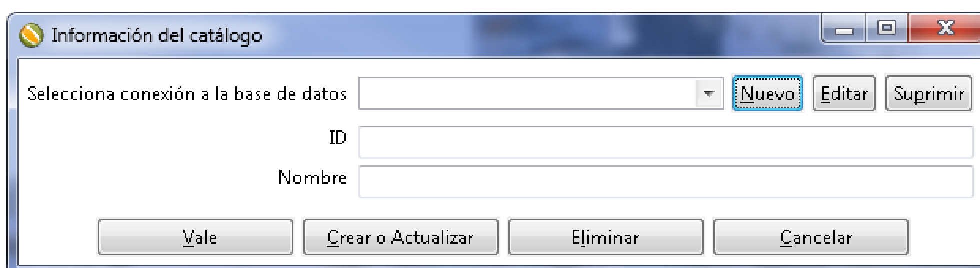
Al no tener ningún repositorio creado aún, se debe proceder a la creación de uno dando clic en el icono verde con el símbolo "+" en la esquina derecha superior de dicha ventana.



Seleccionar tipo de repositorio que se desea utilizar, en este caso un repositorio de base de datos, dando clic en la opción 1 de la lista en la ventana emergente:



Ya seleccionado el tipo de repositorio damos clic en el botón "Vale", aparecerá la siguiente ventana:



En esta se solicita indicar la conexión hacia la base de datos del repositorio, damos clic en el botón “nuevo”, aparecerá la ventana de creación de conexión.

Database Connection
General
Advanced
Options
Pooling
Clustering
Connection Name:
kettle-infhos
Connection Type:
Hadoop Hive
Hypersonic
IBM DB2
Infobright
Informix
Ingres
Ingres VectorWise
Intersystems Cache
KingbaseES
LucidDB
MS Access
MS SQL Server
MS SQL Server (Native)
MaxDB (SAP DB)
MonetDB
MySQL
Access:
Native (JDBC)
ODBC
JNDI
Settings
Host Name:
SERVER
Database Name:
KETTLE
Instance Name:
SQLEXPRESS
Port Number:
1433
User Name:
Xaman
Password:
Use .. to Separate Schema and Table
Probar
ta de funcior
Explorar
OK
Cancel

Buscar en el área de selección “Connection Type” el tipo de conexión MS SQL Server y en el área de selección “Access“ la opción Native JDBC,

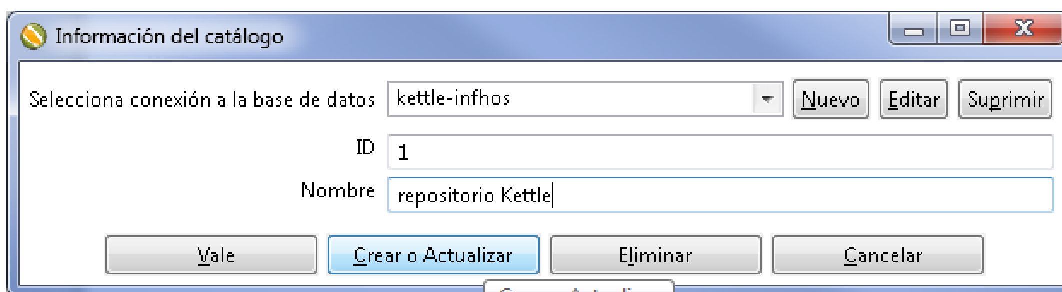
En la sección "Settings" se debe ingresar todas configuraciones para la conexión a la base de datos destinada para el repositorio de KETTLE:

1. Conection Name: kettle-infhos
2. Host Server: SERVER
3. Database Name: KETTLE
4. Insatance Name: SQLEXPRESS
5. Port Number: 1433
6. Username: Xaman
7. Password: Xaman

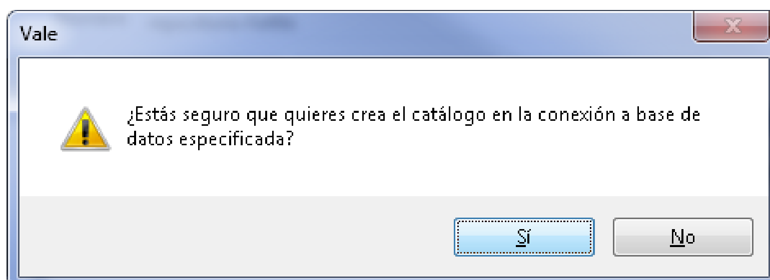
La casilla "Use .. to Separate Schema and Table" debe quedar vacia.

Dar clic en el botón probar para verificar la configuración exitosa de la conexión. Si la conexión es exitosa se da clic en el botón OK

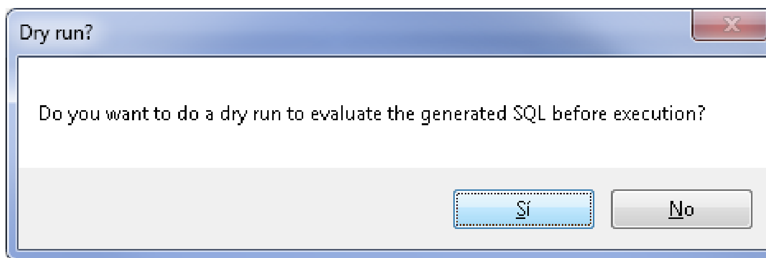
Aparecerá en la ventana anterior, la conexión ya seleccionada. En el campo ID escribir el número 1, en el campo Nombre se debe escribir "repositorio Kettle" y dar clic en el botón "Crear o actualizar" para crear la estructura de la base de datos repositorio de Kettle.



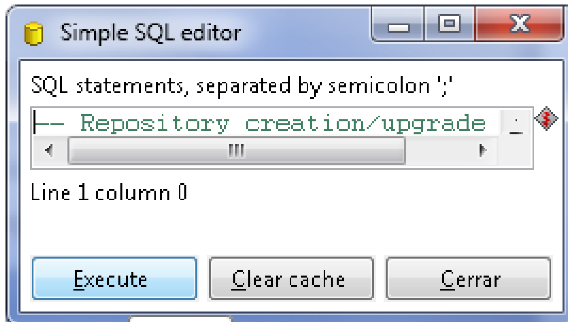
En la ventana emergente de confirmación dar clic en el botón "Si"



Y de nuevo en la ventana emergente clic en "Si"

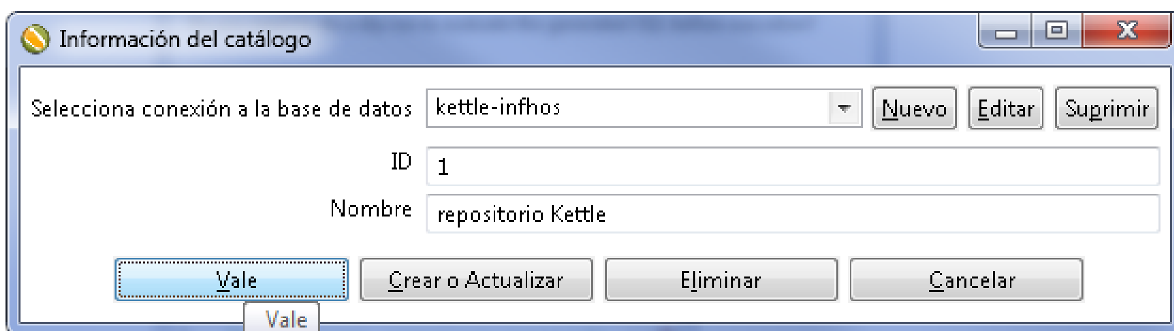


Aparecerá la ventana de ejecución de script. Dar clic en el botón "Execute"



Aparecerá una ventana con el reporte de la ejecución, en esta ventana dar clic en el botón "Vale"

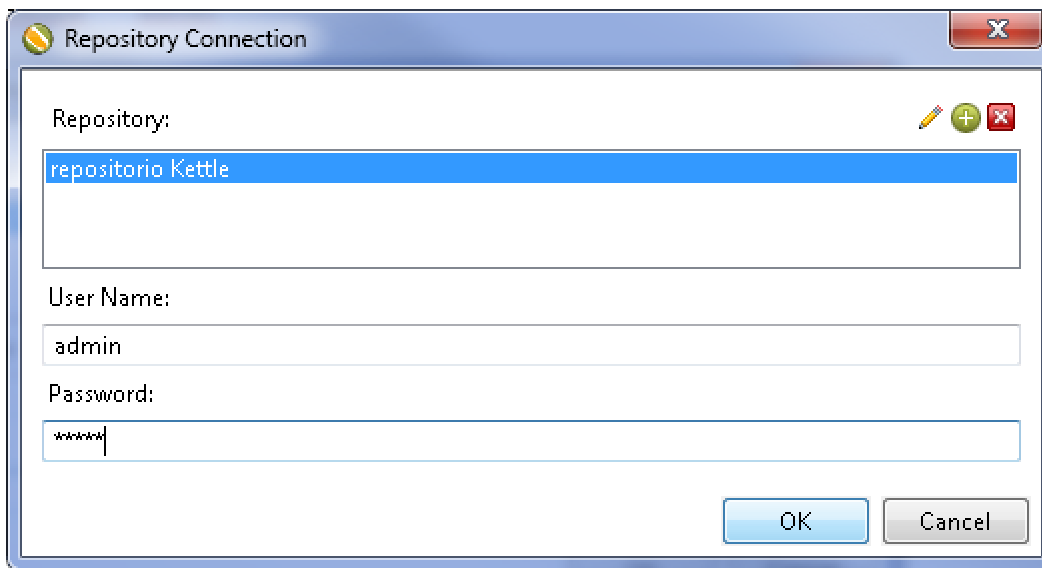
Al finalizar la ejecución, dar clic en el botón "Cerrar" y en la ventana anterior el botón "Vale"



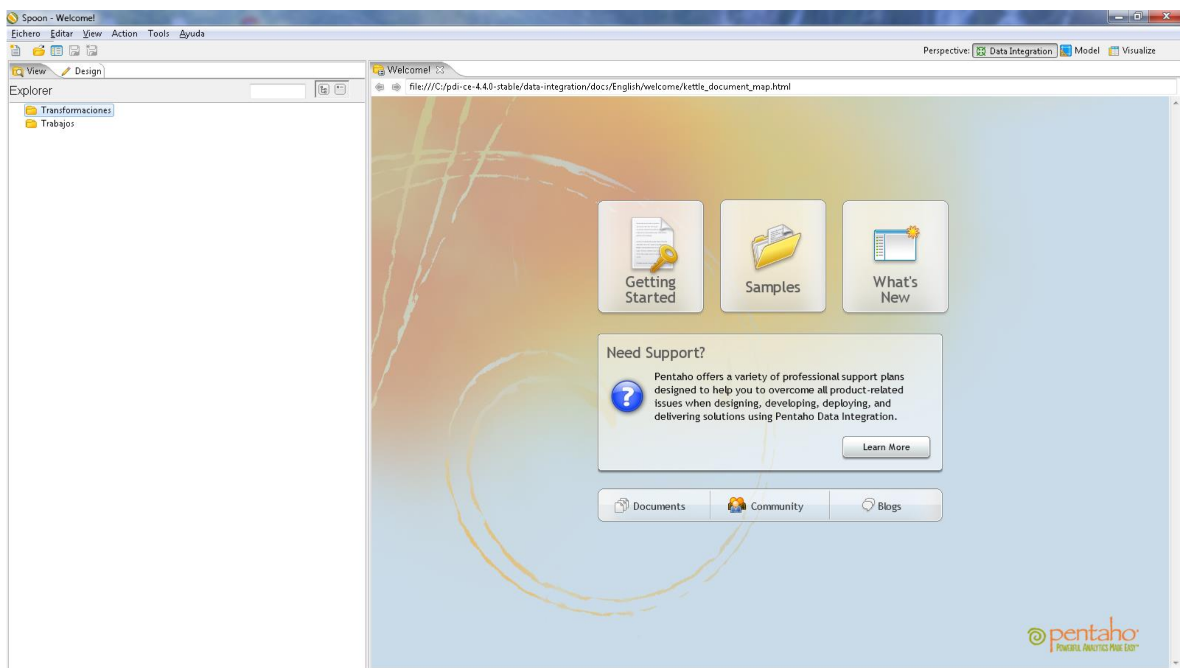
Y en la ventana anterior "Repository Connection", se debe elegir "Repositorio Kettle",

Ingresar los accesos:

1. "User Name" admin
2. "Password" admin



Clic en el botón "OK" para seleccionar el repositorio y ser direccionado a la ventada de diseño de Kettle.

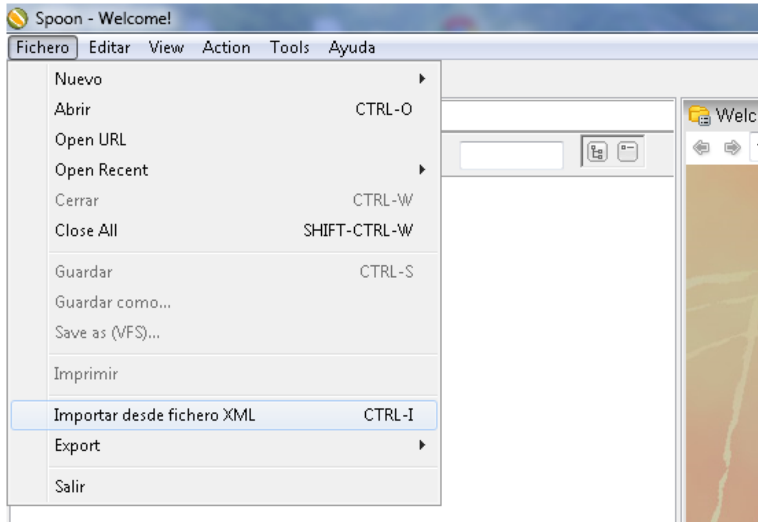


.

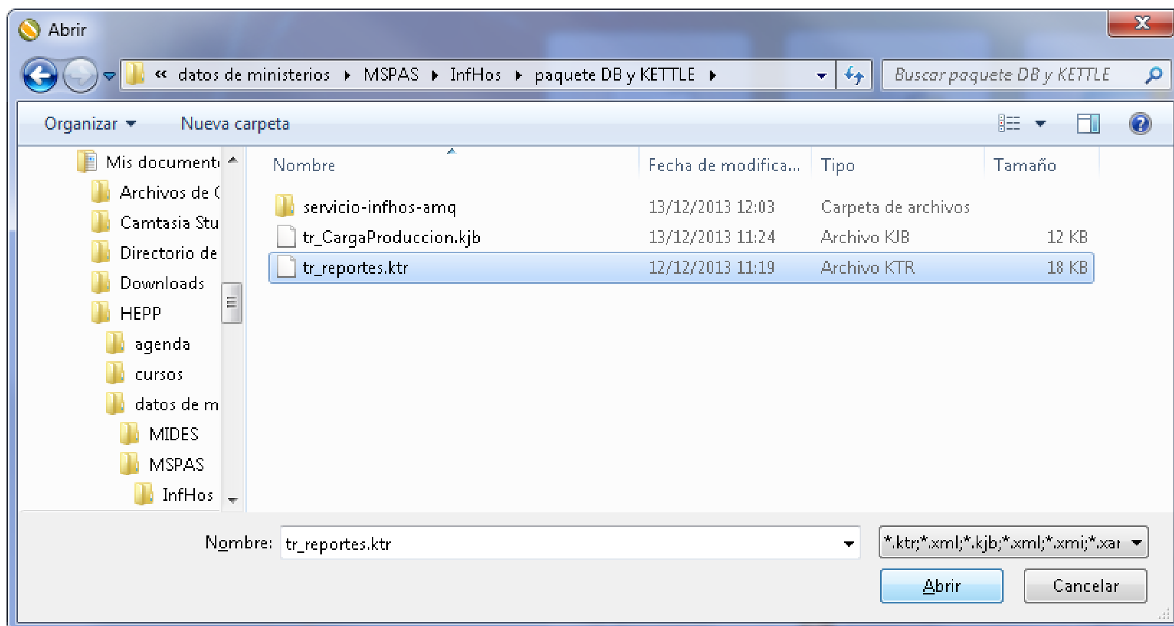
5.2. Importar la transformación y el trabajo prediseñados

En el paquete de recursos descomprimido (recursos.zip) que debería estar alojada en la ruta *C:\ZV\recursos*, se encuentra la transformación y el prediseñados para el proceso.

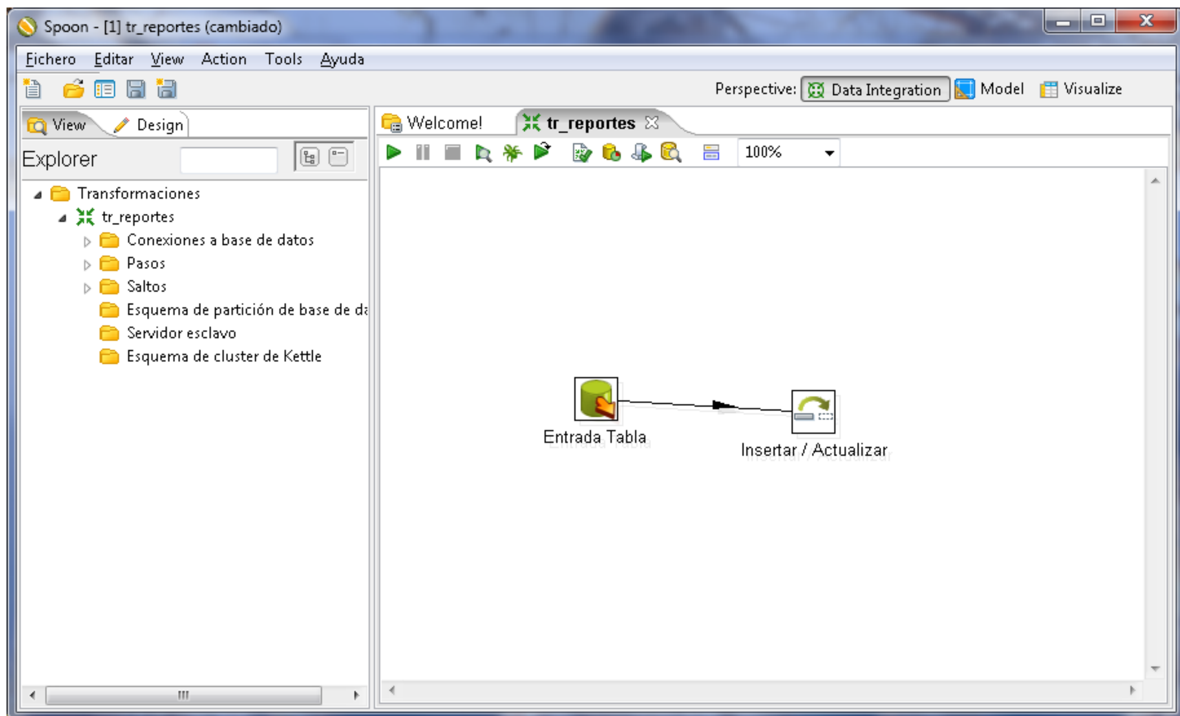
Para importarlos al diseñador actual se debe ir al menú "fichero" en la esquina superior izquierda del diseñador y elegir la opción "importar desde XML":



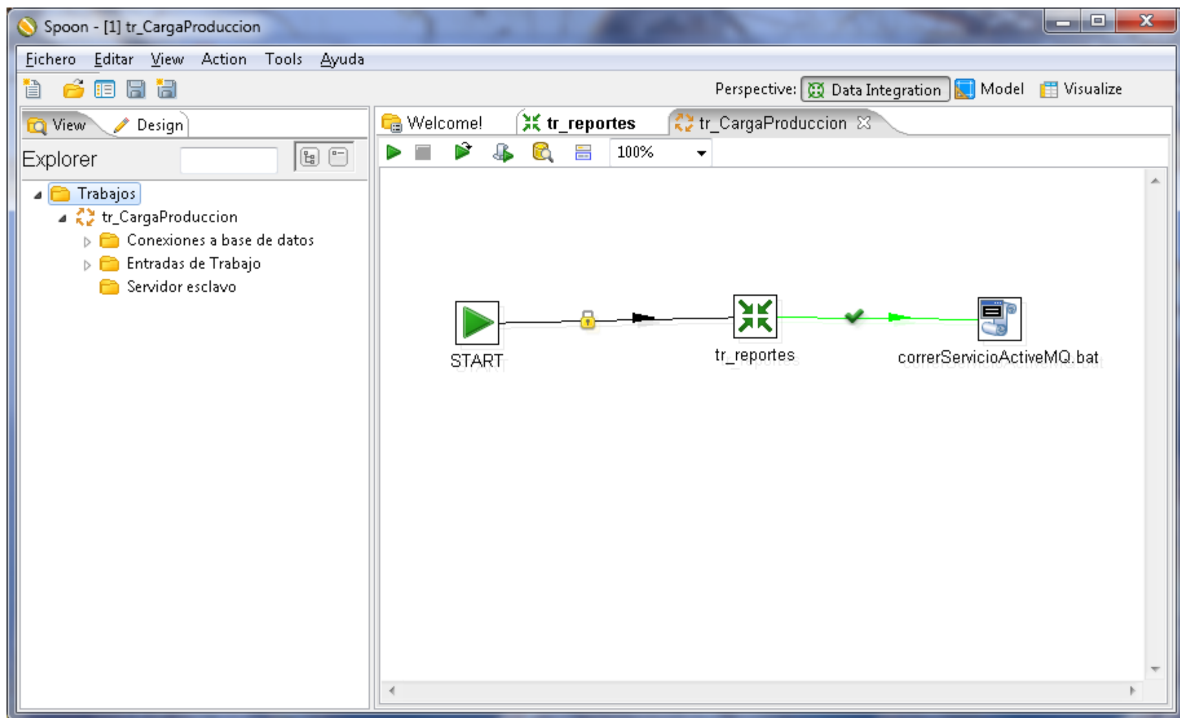
Luego se debe seleccionar de la carpeta de recuros descargada anteriormente el archivo *tr_reportes.ktr*



Y luego clic en abrir. Luego de esto aparecera en la ventana del diseñador la transformación



Se repiten los pasos anteriores, pero ahora para el archivo *tr_CargaProduccion.kjb*



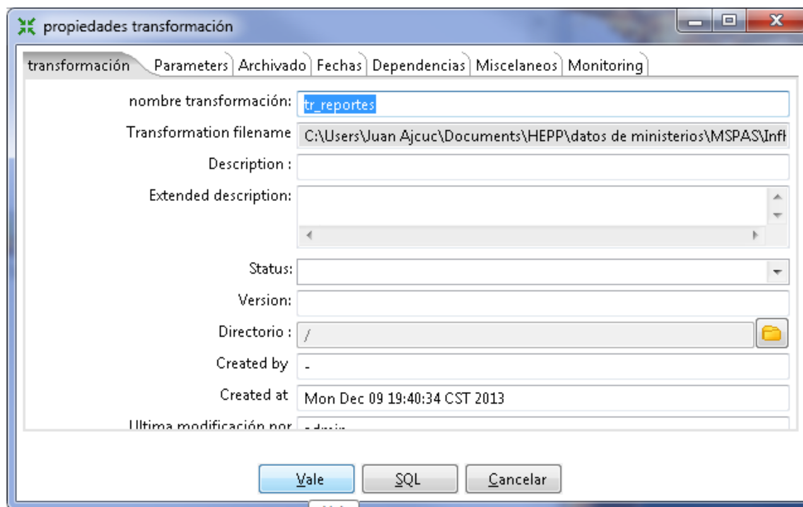
Y aparecera una nueva pestaña con el diseño del trabajo.

5.2.1. Guardar el trabajo y la transformación en el repositorio local

Para guardar los archivos importados en el repositorio local se debe seleccionar la pestaña con nombre "tr_reportes"



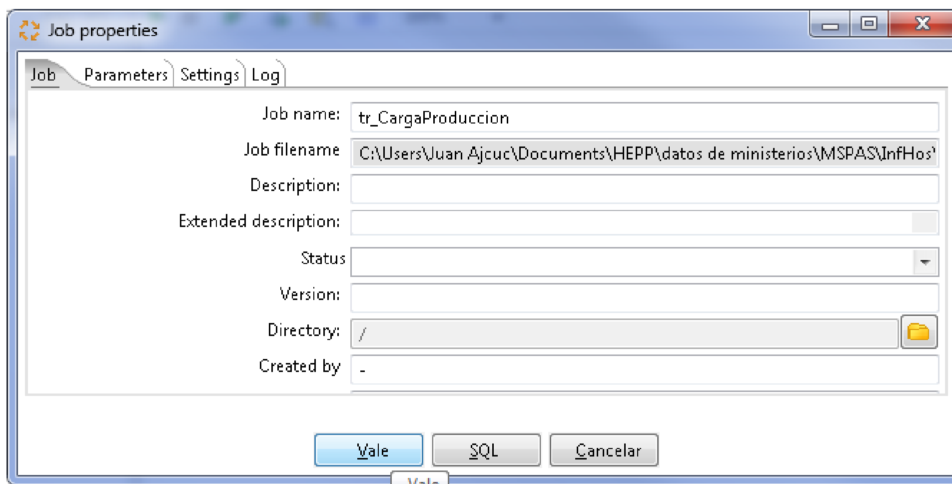
Luego clic en el menú "Fichero" y seleccionar la opción "Guardar", en la ventana emergente se dejan todos los datos sin modificar y se procede a dar clic en el botón "Vale"



Y lo mismo para la pestaña "tr_CargaProduccion"



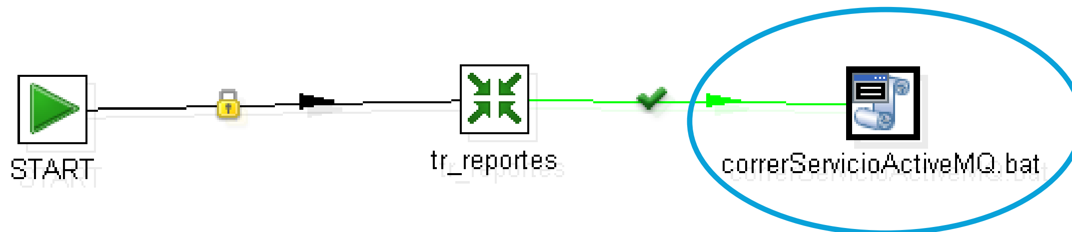
En el menú "fichero" clic en "Guardar" y luego clic en el botón "Vale" en laventana emergente.



5.3. Configuración de la ejecución del servicio productor de datos

Para lograr transferir los datos al ambiente remoto para construcción de los cuadros de mando, se debe configurar el último paso en el diseño del trabajo "tr_CargaProduccion".

En la pestaña "tr_CargaProduccion" localizar el icono "correrServicio ActiveMQ.bat" en el diseño y hacer doble clic sobre el mismo.



Aparecerá una ventana emergente para la configuración del paso, en el campo "Script File name", dar clic en el botón examinar y localizar dentro de la carpeta de recursos descargada anteriormente, que debería estar alojada en la ruta *C:\ZV\recursos*, ingresar a la carpeta *sevicio-infhos-amq* y dentro de la misma el archivo *correrServicioActiveMQ.bat*, luego clic en abrir y en la ventana anterior, clic en el botón "Vale".

Se debe volver a guardar el trabajo.

5.4. Iniciar el servicio de ejecución del ETL

El trabajo se debe ejecutar periódicamente de manera automática.

Para iniciar el servicio y dejarlo ejecutándose de manera recurrente se debe iniciar el mismo dando clic en el botón "iniciar trabajo" en la ventana de diseño del trabajo "tr_CargaProduccion"



Y en la ventana emergente dar clic en el botón ejecutar.

PROYECTO DE POLÍTICAS EN SALUD Y EDUCACIÓN

Ejecutar una transformación

Ejecución local, remota o clustered

Ejecución local

Ejecución remota

Servidor remoto

Pass export to remote server

Details

Habilitar modo seguro

Clear the log before execution

Nivel de registro

Rowlevel (very detailed)

Fecha de Ejecución (yyyy/MM/dd HH:mm:ss)

Starting point of job

Parameters

#	Parameter	Value	Default value
1			

Variables

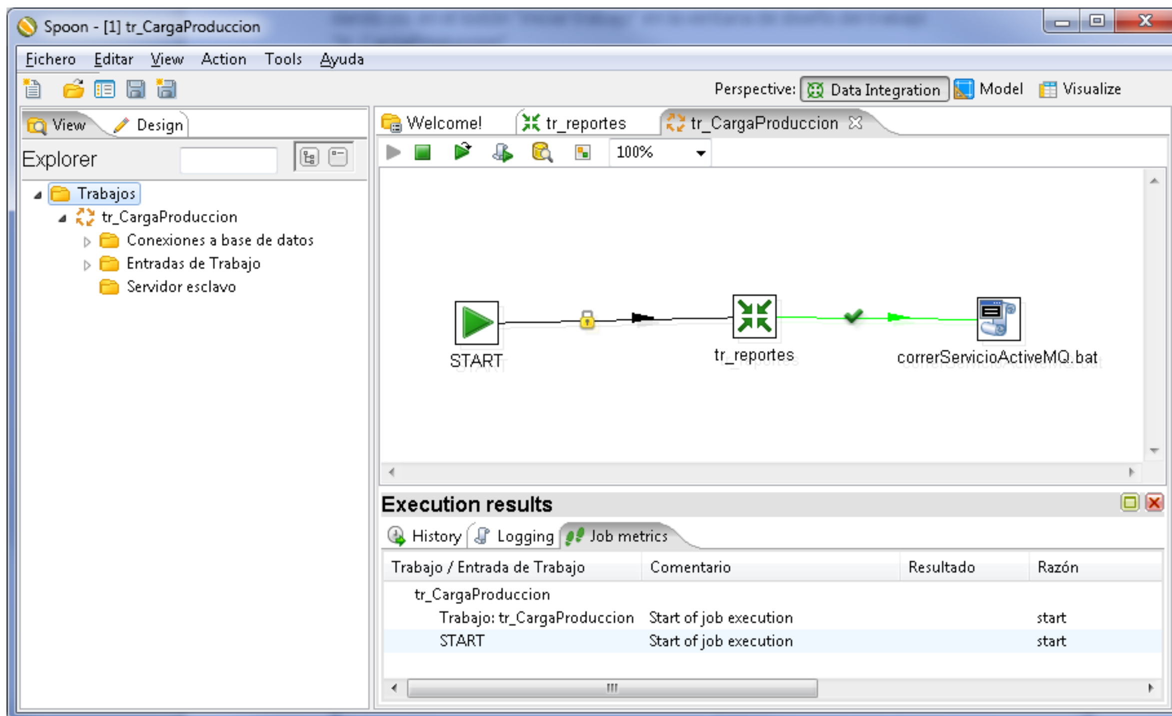
#	Variable	Valor
1		

Parámetros

Ejecutar Cancelar

Ejecutar

El trabajo quedara iniciado y esperando el tiempo de ejecución.



5.4.1. Nota

Para que la ejecución sea efectiva, es necesario que la herramienta de diseño de kettle quede abierta en todo momento.

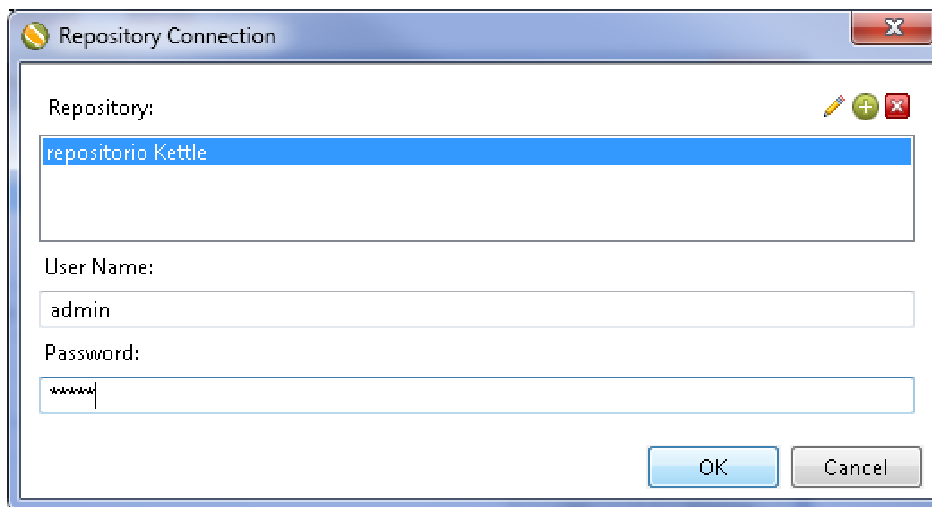
Si esta de llegase a cerrar por alguna razón (reinicio del servidor, cerrado de la ventana, etc.) el proceso no se ejecutará automáticamente.

Para que el proceso de ejecución quede automático de nuevo, simplemente se debe abrir de nuevo la herramienta de diseño de Kettle, abrir el trabajo e iniciar la ejecución del mismo.

Ingresar la carpeta descomprimida *pdi-ce-4.4.0-estable*. Dentro de la misma ingresar a la carpeta *data-integration* y se debe dar doble clic sobre el archivo *spoon.bat*

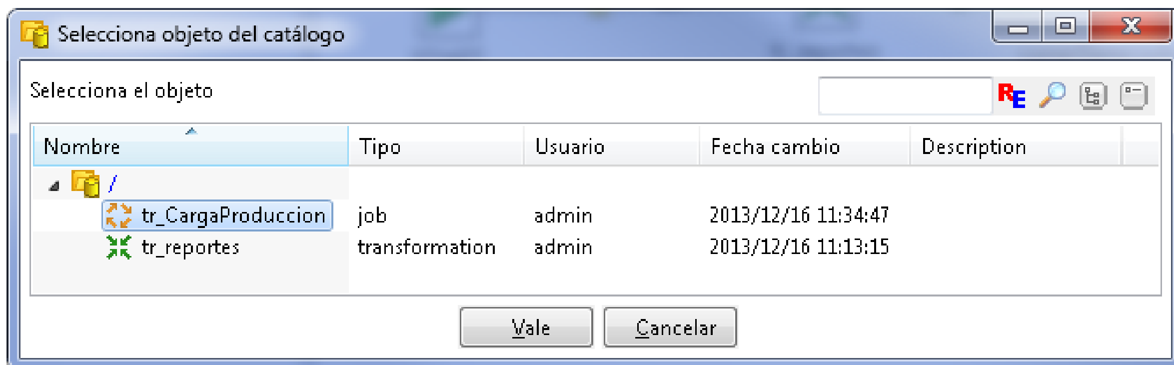
En la ventana de selección de repositorio, elegir "Repositorio Kettle", e ingresar los datos:

1. "User Name" admin
2. "Password" admin



Dar clic en “OK”

Si no aparece el trabajo “tr_CargaProduccion” en la ventana de diseño, ir al menú “fichero2 y a la opción “Abrir”, seleccionarlo de la lista y luego dar clic en el botón “Vale”



Dar clic en el botón “iniciar trabajo” en la ventana de diseño del trabajo “tr_CargaProduccion”



Y en la ventana emergente dar clic en el botón ejecutar.